March 25, 2015

American Housing Survey

Guide to Estimating Variances for the American Housing Survey

U.S. Census Bureau
Department of Commerce

Department of Housing and Urban Development

Table of Contents

This page is intentionally blank

1. Introduction

The American Housing Survey (AHS) provides estimates of many housing statistics in the United States. Standard errors for specific estimates are also provided for the published estimates. Public Use Files (PUFs) are additionally provided to the public; PUFs allow data users to estimate a wide variety of AHS statistics. To estimate the associated variance of those statistics, Generalized Variance Functions (GVFs) and replicate weights are also provided so that data users can estimate the variance. This document is intended to explain how to use the GVFs and replicate weights. As the old saying goes, we could fish for a man and feed him for a day or we could teach him how to fish and feed him for a lifetime. This document teaches you how to fish for your own variances.

Only Design Variances Estimated

By variance, we refer to the sample design variance or simply design variance or the variance from a finite population sample. The variance measured by the GVFs and the replicate weights represents the variance of the estimated statistic if we repeated the sample selection many times and estimated the statistic of interest with each sample. See textbooks by Cochran (1977), Wolter (1984), and Särndal *et al*. (1992) for detailed discussions on sample design and variances.

How the Guide is Organized

The first four sections of the guide describe "how to estimate variances yourself." This part of the guide explains how the variance estimation tools provided by the AHS can be used to estimate variances.

GVFs are provided because they give a simple relationship for the size of the estimated number of housing units relative to the size of the variance. GVFs for the AHS are provided for each survey year, to cover several important domains of interest. All of the GVFs provided are limited to providing variances for totals of housing units.

Replicate weights can be used to estimate sampling variance for any complex statistic from the survey design. The guide provides several examples that show how to use replicate weights to estimate variances of several types of statistics.

The last two sections of the guide explain how the GVFs and the replicate weights are calculated. We provide these two sections for transparency, context, and background for sophisticated data users.

Scope of Guide

The scope of this document includes both the National (AHS-N) and Metro (AHS-MS) samples for 2009 through 2013. How we calculate the replicate weights is different for the National and Metro samples; however, how they are used to estimate variances is the same. Similarly, the National and the Metro GVFs are used in the same way to estimate variances.

We do not provide a review of the sample design for the AHS-N and AHS-MS samples, but refer to the reader to Appendix B of the general publication of AHS estimates for each year.

In 2013, the AHS has a new feature called the split samples. The overall AHS sample for both the AHS-N and AHS-MS was randomly split into two parts, which we refer to as split sample 1 and split sample 2. Sample housing units (HUs) of each split sample were asked all of the regular AHS questions which produce what we will call the "full-sample" variables on the PUF. Additionally, a different set of questions were asked of each of the split sample HUs which we will call the "split-sample" variables on the PUF. For example, split sample 1 was asked about types of public transportation used and split sample 2 was not. Estimates about types of public transportation used, therefore, can only be derived by using split sample 1. Because the items within the split sample modules were only asked of half of the sample, a different set of weights are available for each split sample.

This guide explains how to estimate variances for all data items in AHS – for the full sample, split sample 1, and split sample 2. Generally, the methodology is the same and the difference is that different replicate weights are applied to the three types of variables.

2. How to Estimate Variances with a GVF

Although replicate weights have many advantages over GVFs, some variance estimates from replicate weights can be quite computationally demanding. GVFs are easier to use than replicate weights, they stabilize variance estimates, and are computationally more efficient than variances estimated from replicate weights. Wolter (1985; chapter 5) provides a technical introduction to GVFs and gives additional reasons why GVFs are often used.

For convenience, Appendices B and C of this document include a compilation of prior GVF parameters for the AHS-N and AHS-MS surveys, respectively.

Why GVFs Are Used

GVFs are easy to calculate and since variance estimates are based on sample data they have variances of their own. The estimated variance for a survey estimate generally has less precision than the survey estimate itself. This means that the estimates of variance for the same statistic may vary considerably from year-to-year or for related characteristics in a given year. GVFs provide some stability to the estimates of variance by averaging the variances from estimates of similar size.

How to calculate the variance with a GVF

Let $\hat{N}$ be an estimator of a total number of units (sum of the weights) within a domain of interest and $v(\hat{N})$ be its variance. We can then calculate the variance of $v(\hat{N})$ as a function of $\hat{N}$ as

$$\frac{v(\hat{N})}{\hat{N}^2} = a + \frac{b}{\hat{N}} \tag{2.1}$$

where *a* and *b* are the parameters of the GVF model.

The parameters *a* and *b* of expression (2.1) are provided each year for estimates of different domains. The left side of (2.1) is the relative variance or relvar, i.e.,

$$relvar(\hat{N}) = \hat{v}(\hat{N}) / \hat{N}^2 . \tag{2.2}$$

The coefficient of variation (CV) is defined as the square root of the relvar, i.e.,

$$cv(\hat{N}) = \sqrt{\hat{v}(\hat{N})} / \hat{N} . \tag{2.3}$$

The domain GVFs can be used to produce variances for subsets of the data based on geographic, demographic, and other characteristics.

Example 2.1 Estimating the Variance of a Total

This example shows how to estimate the overall variance for the estimated number of owner-occupied housing units in the U.S. during 2013 using a GVF. The first step is to get the GVF parameters and the estimate of the statistic of interest; the number of 2013 owner-occupied housing units in the U.S. In 2013, there were 75,650,326 owner-occupied housing units in the U.S.

Next, we get the appropriate GVF parameters for 2013. Table 2.1 provides the parameters for the GVF of the 2013 AHS estimates.

Table 2.1: 2013 GVF Parameter Set (Full Sample)

GVF	*a*	*b*[*]
Overall Total Housing Units Estimates	-0.000050	6.68
Owner-Occupied Domain Estimates	-0.000077	7.03
Renter-Occupied Domain Estimates	-0.000073	3.65

* The parameter *b* is represented in thousands.

Inserting the estimates of owner-occupied and the parameters from Table 2.1 into expression (2.1), we get the relative variance:

$$\frac{v\left(\hat{N}\right)}{\hat{N}^2} = \frac{a \cdot \hat{N}^2 + b \cdot \hat{N}}{\hat{N}^2} = \frac{-0.000077 \cdot 75{,}650.326^2 + 7.03 \cdot 75{,}650.326}{75{,}650.326^2} = \frac{91{,}152.96}{75{,}650.326^2} = (0.00399)^2. \tag{2.4}$$

The computation yields a coefficient of variation (CV) of 0.00399. The estimated variance $v\left(\hat{N}\right)$ is in thousands and yields 91,152,960 with standard error of 301,915.

GVFs of the Appendices

Appendices B and C include the 2013 and prior GVFs for the national and metro samples, respectively. Note that the GVFs work with estimates in the thousands so variances must be multiplied by $1{,}000^2$ and standard errors must be multiplied by 1,000 after applying (2.1).

Limitations of GVFs

The GVF parameters are estimated from models used to estimate the variance of estimated totals of households. As with any model, estimates should only be applied to GVFs that are within the range of values used to estimate the parameters. Since the GVFs only used totals of HUs, the GVFs are only appropriate for estimating totals of HUs. The GVFs provided should not be used to estimate variances of:

- Statistics other than the estimated number of HUs.
- Estimates that involve sample HUs from other domains of interest.

For example, the GVFs should not be used to estimate the variances for estimates of the number of people who rent. The rural GVF should not be used if the estimate involves urban and rural sample HUs.

3. How to Estimate Variances with Replicate Weights

The variance of any survey estimate based on a probability sample may be estimated by the method of replication. This method requires that the sample selection, the collection of data, and the estimation procedures be independently carried through (replicated) several times. Each time the sample is replicated, a different set of estimates is calculated. The dispersion of the resulting estimates then can be used to measure the variance of the full sample.

However, we would not consider repeating any large survey, such as the AHS-N, several times to obtain variance estimates. A practical alternative is to manipulate the full sample several times by applying different weighting factors to the sample units. The manipulation of the weights causes the sample data to represent a different number of housing units in each replicated sample. We then apply the estimation procedures (e.g., mean, median, sum, etc.) to these weighted random samples. We refer to these random samples as replicate samples or simply replicates. For the AHS-N, we used 160 replicates to calculate the AHS-N variance estimates.

The replicate weights should only be used in creating variances and should not be used to create independent estimates.

The user should also note that the replicate weights are calculated using information from the sample. Therefore, the 2011 AHS-N replicate weights are applicable for use on only 2011 AHS-N data. Replicate weights for 2013 will be provided with the 2013 public use file.

Use of Replicate Estimates in Variance Calculations

Calculate variance estimates using Fay's Balanced Repeated Replication (BRR) method (Judkins 1990) for AHS-N and AHS-MS estimates using the following formula:

$$\hat{v}(\hat{\theta}) = \frac{4}{160}\sum_{r=1}^{160}\left(\hat{\theta}_r - \hat{\theta}\right)^2 \tag{3.1}$$

where $\hat{\theta}$ is the weighted estimate of the statistic of interest; such as a total, median, mean, proportion, regression coefficient, or log-odds ratio, using the weight for the full sample and $\hat{\theta}_r$ is the replicate estimate for replicate *r* of the same statistic using the replicate weights. To calculate a standard error, the measure of dispersion when parameter estimates are calculated through repeated sampling from the population, obtain the square root of the variance estimate.

The value of 4 in equation (3.1) arises from the use of successive difference replication (SDR). See Fay and Train (1991) and Ash (2014) for more background on SDR.

To ensure confidentiality of the data, some characteristics have either been bottom coded or top coded. This procedure places a lower (or upper) boundary on the published value for the variable in question. Therefore, some estimates calculated from the Public Use File may differ from the estimates provided in the AHS-N publication tables.

Using Replication to Estimate Variances

The following example illustrates how a statistic would be estimated, replicated, and combined to form a variance estimate. We are going to estimate the variance using the 160 replicate weights provided for the AHS.

Please note that in the following example, the weights in Replicate 0 equal the Full-Sample Weight. The Hadamard matrix was used to derive replicate factors to apply to individual Full-Sample weights in creating replicate weights.

Example 3.1 Estimating the Variance of the Total Number of Housing Units in a Domain

The goal of this example is to estimate the total number of owner-occupied housing units in the U.S. for 2013 and its corresponding estimate of variance. Assume we have five sample cases with responses shown below when asked if they owned their house (tenure) during the time of interview.

Table 3.1: Example of Estimating Variances with Replication

Sample HU	Owned House?	Full-Sample Weight	Replicate Weights				
			Replicate 1	Replicate 2	Replicate 3		Replicate 160
1	**YES**	**2,995**	**2,646**	**3,594**	**810**	…	**1,006**
2	NO	2,359	1,403	1,358	1,317	…	1,415
3	**YES**	**1,509**	**846**	**1,106**	**810**	….	**791**
4	**YES**	**2,283**	**1,360**	**1,349**	**1,271**	…	**1,269**
⋮	⋮	⋮	⋮	⋮	⋮		⋮
n	**YES**	**2,497**	**1,475**	**1,414**	**1,468**	…	**1,385**

In AHS, the full-sample estimate and the full-sample weight are referred to as the replicate estimate 0 and replicate weight 0, respectively.

Step 1: Calculate the full-sample weighted survey estimate.

The statistic of interest is the total number of owner-occupied housing units in the U.S. for 2013. Add the full-sample weights of the sample cases that responded “YES” to the tenure question. Therefore, the total number of owner-occupied housing units survey estimate is calculated as follows:

Full-Sample Owner-Occupied Estimate $\hat{N}$ = 2,995 + 1,509 + … + 2,497 = 75,650,326

Step 2: Calculate the weighted survey estimate for each of the replicate samples.

The replicate survey estimates are as follows:

Replicate 1 Owner-Occupied Estimate $\hat{N}_{r=1}$ = 2,646 + 846 + … + 1,475 = 75,709,529
Replicate 2 Owner-Occupied Estimate $\hat{N}_{r=2}$ = 3,594 + 1,106 + … + 1,414 = 75,655,411
Replicate 3 Owner-Occupied Estimate $\hat{N}_{r=3}$ = 810 + 810 + … + 1,468 = 75,630,527
⋮ ⋮
Replicate 160 Owner-Occupied Estimate $\hat{N}_{r=160}$ = 1,006 + 791 + … + 1,385 = 75,681,238

Step 3: Use these survey estimates in formula (2.1) to calculate the variance estimate for the total owner-occupied population.

$$
\begin{aligned}
\hat{v}(\hat{N}) &= \frac{4}{160}\sum_{i=1}^{k}\left(\hat{N}_r - \hat{N}\right)^2 \\
&= \frac{4}{160}\times\begin{bmatrix}(75{,}709{,}529-75{,}650{,}326)^2 + (75{,}655{,}411-75{,}650{,}326)^2 \\ +(75{,}630{,}527-75{,}650{,}326)^2 + \ldots + (75{,}681{,}238-75{,}650{,}326)^2\end{bmatrix} \\
&= 0.025\times\left[3{,}504{,}995{,}209 + 25{,}857{,}225 + 392{,}000{,}401 + \ldots + 955{,}551{,}744\right] \\
&= 94{,}853{,}015{,}615.
\end{aligned}
$$

The estimate of the variance of total owner-occupied population is $\hat{v}(\hat{N}) = 94{,}853{,}015{,}615$.

The survey estimate for total owner-occupied population is 75,709,529 housing units. This survey estimate has an estimated variance of 94,853,015,615, or a standard error of 307,982 housing units.

The three steps of Example 3.1 will be used throughout the guide to calculate variances. Sometimes Steps 1 and 2 will be combined since estimating the statistic with the regular weight (or for replicate 0) can be done with estimating the statistic with the 160 replicate weights.

Confidence Intervals and Significance Tests

Once the standard error is calculated, it can be used to create confidence intervals and perform significance tests. Use the estimate where the equation requires the standard error. For means, medians, totals, and regression coefficients, the degrees of freedom will equal the number of replicates. For detailed examples of proportions and log-odds ratios (as well as the other statistics given here), refer to the unabridged version of this document. More discussion of estimating confidence intervals is included in section 5.

AHS-N Replicate Weight File Description

To access the replicate weight file, go to the Census Bureau's AHS website at: http://www.census.gov/programs-surveys/ahs/data.html. Once navigation to the site is completed, select the "2013" tab and select the AHS 2013 National Public Use File (PUF). The file repwgt.sas7bdat contains the replicate weights and match key required to merge the replicate weight file to the public use survey data files. This is a SAS formatted dataset of 84,355 records. The match key is the variable CONTROL (a character variable with length 12). When reading this file into SAS, the following example will provide the replicate weight data set (Input-related items are highlighted in yellow).

Figure 3.1.1: SAS Code for Reading in the Replicate Weights

```
libname puf2011 'C:\directory_name_where_the_2011_files_are_located';
libname puf2013 'C:\directory_name_where_the_2013_files_are_located';

data repwgt2013;
      set puf2013.repwgt;
      keep control repwgt0-repwgt160;
run;

data repwgt2011;
      set puf2011.repwgt;
      keep control repwgt0-repwgt160;
run;
```

This SAS code identifies the variables in the dataset for the full sample. Split sample weights are also included on the file but have been omitted from the KEEP statement for the purpose of this guide and its examples. The first variable is REPWGT0, the second is REPWGT1, and so on through REPWGT160, and the last variable is CONTROL. The replicate weights are in REPWGT1-REPWGT160. The full sample weight is in REPWGT0.

The REPWGT file and the public use survey data file both have the full sample weight. On the REPWGT file the variable name is REPWGT0, but on the public use survey data file the variable name is WGT90GEO. The full weight on the REPWGT file can be used as means of verifying that the files are properly merged to the public use survey data.

Merging the AHS-N Replicate Weight File with the AHS-N Public Use File

To use the replicate weights with the PUF, first merge the two files together. To do this, obtain:

AHS-N Public Use File (file name = NEWHOUSE)
AHS-N Replicate Weight File (file name = REPWGT)

Merge the two files using CONTROL. Before merging, however, ensure that both files are sorted by CONTROL. Both files will have the same number of records because both files will have the same sample housing units.

Figure 3.1.2: SAS Code for Combining AHS data and replicate weights

```
data data2011;
  merge puf2011.NEWHOUSE repwgt2011;
    by control;
run;

data data2013;
  merge puf2013.NEWHOUSE repwgt2013;
    by control;
run;
```

The data files called data2011 and data2013 in Figure 3.1.2 will be used throughout this document.

Statistical Software for Calculating Variances

There are several statistical software packages that employ replicate weights and equation (3.1). We now review a few of the software packages.

R Software. Documentation for package 'survey' version 3.6-12, User Guides and Package Vignettes (http://rss.acs.unt.edu/Rdoc/library/survey/html/00Index.html)

Stata. A good review of the use of STATA is given by Kreuter and Valliant (2007).

Wesvar. This software is developed and distributed by Westat and can use replicate weights directly to estimate variances of a complex survey design.

SAS. Within SAS, there are currently four specialized "survey" procedures that use replicate weights directly. Table reviews the procedures.

Table 3.2: Summary of SAS "SURVEY" PROCs

PROC name	Can be used to…
SURVEYMEANS	Calculate basic statistics
SURVEYFREQ	Completes categorical data analysis
SURVEYREG	Completes regression analysis
SURVEYLOGISTIC	Completes logistic regression analysis

The standard errors calculated by the above procedures need to be multiplied by 2. This is done because of a nuance in the SAS procedures. Recall equation (3.1) which calculated variance estimates using Fay's Balanced Repeated Replication (BRR):

$\hat{v}(\hat{\theta}) = \frac{4}{160}\sum_{r=1}^{160}(\hat{\theta}_r - \hat{\theta})^2$. The SAS procedures above do not include the 4 used in the calculation. As a result, to correctly estimate the variance, one must multiply the calculated variance by 4 or the calculated standard error by the square root of 4.

Mukhopadhyay *et al*. (2008) provides an excellent review of replication-based variances methods and the "survey" procedures of SAS.

Each of the procedures of Table 3.2 can use the replicate weights in its analysis. For example, the surveymeans procedure and the replicate weight file can be used together to generate a standard error for a population total estimate. Input related items are highlighted in yellow below.

Figure 3.1.3: SAS Code Using PROC SURVEYMEANS

```
proc surveymeans data=data2013 sum std clsum cvsum varmethod=brr;
  var variable;
  weight repwgt0;
  repweights repwgt1-repwgt160;
run;
```

PROC SURVEYMEANS does not calculate BRR variance estimates for medians. Support for the use of BRR to estimate the sampling variance of a median can be found in Thompson and Sigman (2000).

Since we are most familiar with SAS, the next section will provide examples using SAS. It is recommended that if the reader is using a software package other than SAS, he/she should refer to their own software's documentation or customer support for details on the use of replicate weights in variance estimation. However, the examples of the next section may facilitate an understanding of how to create the replicate variance estimates in their own software.

Example 3.2 Regression model with PROC SURVEYMEANS

In this example, we estimate the mean number of persons within households. The output for PROC SURVEYMEANS will also include the estimated variance of the mean number of persons. Figure 3.2 provides the SAS code for PROC SURVEYMEANS.

Figure 3.2.1 provides the SAS code that can be used to estimate the mean and total number of persons from the 2013 AHS-N.

Figure 3.2.1: Example of SAS Code Using PROC SURVEYMEANS

```
proc surveymeans data=data2013 mean sum varmethod=brr;
  var per;
  weight repwgt0;
  repweights repwgt1-repwgt160;
  ods output statistics= mean1;
run;
data mean2;
      set mean1;
            *Remember to multiply standard error by 2;
            se = stderr*2;
            var = se**2;
      keep varname mean se var;
run;
proc print data = mean2;
run;
```

The SAS code of Figure 3.2.1 generates the output seen in Figure 3.2.2.

Figure 3.2.2: SAS Output for PROC SURVEYMEANS

VarName	Mean	se	var
PER	2.500446	.006418829	.000041201

PROC SURVEYMEANS calculated the estimated number of persons in all HUs as 289,680,753 with a standard error of 1,386,088. The estimate of the mean number of persons in each HU is 2.5 with a standard error of 0.006.

4. Examples of Calculating Variances with SAS

This section includes examples of calculating the variance of a total, median, and mean using replicate weights.

Example 4.1 Estimating the Variance of a Total

In example 2.1, we showed how to estimate the variance of the estimated number of owner-occupied housing units in the U.S. for 2013 with a GVF. In example 3.1, we generally demonstrated how to estimate the variance for the same statistic using replicate weights. We now demonstrate how to estimate the same total directly with the replicate weights.

First, read in the data file and keep those sample units in the domain of interest. For our example, the domain of interest is all owner-occupied housing units in the U.S. in 2013. In figure 4.1.1, we start by keeping only those sample HUs of interest; the 2013 completed interviews who are owners.

Figure 4.1.1: SAS Code for Combining AHS Data and Replicate Weights

```
/* Flag the data records desired after creating the SAS data sets.
   This example flags owner occupied housing units                 */

data data1;
  set data2013;
     where (STATUS = '1' and TENURE = '1');
run;
```

Next, Figure 4.1.2 shows how we can calculate the full-sample and replicate estimates of the number of owner occupied housing units in the U.S. in 2013. We use the full-sample weights because owner-occupied status was determined of all HUs in the interview and is a full-sample variable.

Figure 4.1.2: SAS Code for Estimating the Variance of Totals

```
/* Steps 1 & 2: Sum the full sample and the 160
   replicate weights and writes them out to a file */

proc means data=data1 sum noprint;
  *The full sample and the replicates.;
  var repwgt0-repwgt160;
  output out=data2 sum=est rw1-rw160;
run;
```

The final step, represented by Figure 4.1.3, is to apply equation (3.1) to the full-sample and replicate estimates.

Figure 4.1.3: SAS Code that Calculates the Estimated Variance of a Total

```
/* Step 3: Use the estimates of the full sample and the
   160 replicates to compute the estimated replicate
   variance(s) using the formula(s) for 160 replicates. */

data data3 (keep = est var se);
    set data2 end=eof ;
    *Fill array with the replicate sums;
    array repwts{160} rw1-rw160 ;
    *Fill array with the squared diffs;
    array sdiffsq{160} sdiffsq1-sdiffsq160;
      do j = 1 to 160 ;
        sdiffsq{j} = (repwts{j} - est)**2 ;
      end ;
        *Sum the squared diffs;
         totdiff = sum(of sdiffsq1-sdiffsq160);
         var = (4/160) * totdiff;
         se = (var)**(0.5);
     output;
run;

proc print data=data3 noobs;
    var est var se;
    format est se comma12.0;
run;
```

The SAS code of Figure 4.1.1-4.1.3 generates the output of Figure 4.1.4.

Figure 4.1.4: SAS Output of Variance of Total

est	var	se
75,650,326	94,853,015,615	307,982

In this example and Example 2.1, the estimate of the number of owner-occupied HUs in 2013 is 75,650,326. However, the estimated standard errors are different: 30,192 from Example 2.1 and 307,982 from Example 4.1. The differing estimates are ok because they are derived from different methods.

In example 4.1, we showed how to estimate the variance of a total. Next, in example 4.2, we will show how to estimate the variance of a mean. The variable of interest will be the mean number of people per household.

Example 4.2 Estimating the Variance of a Mean

An estimated ratio from a survey, defined as $\hat{R} = \hat{Y} / \hat{X}$, is a nonlinear statistic of two estimated totals $\hat{Y}$ and $\hat{X}$. BRR was defined for just this. To estimate the variance of a mean $\hat{\bar{Y}} = \left(\sum \hat{Y}\right) / \hat{N}$, we note that a mean is a special case of the ratio estimator.

This example demonstrates how to estimate the variance of a mean. The specific statistic of interest is the mean person in household for owner-occupied housing units in the U.S. for 2013. The first step of calculating the variance of a mean with replication is to calculate the mean for each replicate. Figure 4.2.1 shows how this can be done.

Figure 4.2.1: SAS Code for Estimating the Variance of a Mean

```
* This empty data set is produced for the merge later. ;
data data4;
     length mean0-mean160 8.;
run;

* Steps 1 & 2: Estimate the replicate estimates for
  replicates 0 to 160.   ;
%macro repss(rep);
proc means data=data1 mean noprint ;
      weight repwgt&rep.;
      var per;
      output out=datal&rep. mean=mean&rep.;
run;
data data4;
      merge data4 datal&rep.;
run;
%mend repss;

%macro doit;
      %do i=0 %to 160;
            %repss(&i.);
      %end;
%mend doit;

%doit;
```

Finally, we apply equation (3.1) to calculate the replicate variance. Figure 4.2.2 shows how this can be done with SAS.

Figure 4.2.2: SAS Code for Estimating the Variance of a Mean

```
* Apply Step 3 to the replicate estimates and estimate the variance.;
data data5 (keep=mean0 var se);
   set data4 end=eof;
      *Fill array with the replicate means ;
       array mean{160} mean1-mean160 ;
       *Fill array with the squared diffs. ;
       array sdiffsq{160} sdiffsq1-sdiffsq160;
          do j = 1 to 160 ;
            sdiffsq{j} = (mean{j} - mean0)**2 ;
          end;
          *Sum the squared diffs. ;
          totdiff = sum(of sdiffsq1-sdiffsq160);
          var = (4/160) * totdiff;
          se = (var)**(0.5);
    output;
run;
proc print data=data5 noobs ;
     var mean0 var se ;
     format mean0 se 8.4 ;
run;
```

The execution of the SAS code of Figures 4.2.1-4.2.2 generates the output seen in Figure 4.2.3.

Figure 4.2.3: SAS Output for Estimating the Variance of a Mean

mean0	var	se
2.5629	.000052790	0.0073

Example 4.3 Estimating the Variance of a Median

This example demonstrates how to estimate the variance of a median. The specific statistic of interest is the median housing value for owner-occupied housing units in the U.S. for 2013.

Estimating the variance of a median is generally the same as a median in that we need to calculate the median for every replicate and then apply equation (3.1). The first and second steps are to calculate the median for each replicate: replicate 0 and replicates 1 to 160. The third step is to apply equation (3.1) to the full-sample and replicate estimates. Figure 4.3.1 shows how this can be done with SAS.

Figure 4.3.1: SAS Code for Estimating the 2013 Median Housing Value of Owner-Occupied HUs

```
* This empty data set is produced for the merge later. ;
data data6;
      length med0-med160 8.;
run;

* Steps 1 & 2: Estimate the replicate estimates for replicates 0 to 160. ;
%macro repss(rep);
proc means data=data1 median noprint ;
      weight repwgt&rep.;
      var VALUE;
      output out=datal&rep. median=med&rep.;
run;
data data6;
      merge data6 datal&rep.;
run;
%mend repss;
%macro doit;
      %do i=0 %to 160;
            %repss(&i);
      %end;
%mend doit;
%doit;

* Apply Step 3 to the replicate estimates and estimate the variance. ;
data data7 (keep=med0 var se);
     set data6 end=eof;
        *Fill array with the replicate means ;
        array med{160} med1-med160 ;
        *Fill array with the squared diffs. ;
        array sdiffsq{160} sdiffsq1-sdiffsq160;
           do j = 1 to 160 ;
             sdiffsq{j} = (med{j} - med0)**2 ;
           end;
            *Sum the squared diffs. ;
            totdiff = sum(of sdiffsq1-sdiffsq160);
            var = (4/160) * totdiff;
            se = (var)**(0.5);
     output;
run;

proc print data=data7 noobs ;
    var med0 var se ;
run;
```

The execution of the SAS code of Figure 4.3.1 generates the output seen in Figure 4.3.2.

Figure 4.3.2: SAS Output for Estimating 2013 Median Housing Costs of Owner-Occupied HUs

med0	var	se
160,000	42,500,000	6,519

Example 4.4 Estimating the Variance of a Regression Parameter Estimate

Correctly estimating the variance of a regression coefficient can be complicated depending on what procedure is used. This example will show a regression model modeling the 2013 value of an owner-occupied housing unit using the number of bedrooms reported in the housing unit. The regression model will first be shown using PROC SURVEYREG with correct parameter estimates and variances. Next, the regression model will be shown using PROC REG with the correct parameter estimates but will generate incorrect variances. Last, how to calculate the correct variance using a variance equation will be demonstrated.

First, PROC SURVEYREG will be used to generate the correct parameter estimates and variances. Figure 4.4.1 shows how this can be done.

Figure 4.4.1: SAS Code for Estimating the Variance of a Regression Parameter

```
proc surveyreg data=data1 varmethod=brr;
      model value = bedrms / solution;
      weight repwgt0;
      repweights repwgt1-repwgt160;
      ods output parameterEstimates = MyParmEst;
run;

data MyParmEstfin;
      set MyParmEst;
            if _n_ = 2;
            se = stderr*2;
            var = se**2;
      drop parameter dendf tvalue probt stderr;
run;

proc print data=MyParmEstFin noobs;
      format estimate se var 8.4;
run;
```

The execution of the SAS code of Figure 4.4.2 produces the output in Figure 4.4.1.

Figure 4.4.2: SAS Output for Estimating Regression Parameters

Estimate	se	var
84,952	2,413	5,824,204

Using the WEIGHT statement in SAS is not a shortcut. We now show how PROC REG and the full sample weights estimate the correct parameter estimates but understate the variances because it does not correctly account for the two-stage sample design of AHS-N or the systematic random sample of both AHS-N and AHS-MS. Figure 4.4.3 shows how this can be done.

Figure 4.4.3: SAS Code for Estimating Regression Parameters

```
* Weighted -- Using the sample design weights.;

proc reg data=data1;
      model value = bedrms;
      weight repwgt0;
      ods output parameterEstimates = MyParmEstw;
run;
data MyParmEstwFin;
      set MyParmEstw;
      if _n_ = 2;
      drop model dependent variable df tvalue probt label;
run;
proc print data=MyParmEstwFin;
run;
```

The execution of the SAS code of 4.4.3 generates the output seen in Figure 4.4.4.

Figure 4.4.4: SAS Output for Estimating Regression Parameters

Estimate	StdErr
84,952	1,586

As expected, the estimate of the regression parameter in Figures 4.4.4 and 4.4.2 agree, but the standard error using PROC REG and the WEIGHT statement under estimates the standard error by 34% or $\left(1-\frac{1{,}586}{2{,}413}\right)$.

Next, we show the worst possible case. We show what happens when PROC REG is used without full sample weights to generate an unweighted modeled parameter estimate with its corresponding standard error. Figure 4.4.5 shows how this can be done.

Figure 4.4.5: SAS Code for Estimating Regression Parameters

```
*Unweighted – Not using the sample design weights.;

proc reg data=data1;
    model value = bedrms;
    ods output parameterEstimates = MyParmEstu;
run;
data MyParmEstuFin;
      set MyParmEstu;
      if _n_ = 2;
      drop model dependent variable df tvalue probt label;
run;
proc print data = MyParmEstuFin;
run;
```

The execution of the SAS code of Figure 4.4.5 generates the output seen in Figure 4.4.6.

Figure 4.4.6: SAS Output for Estimating Regression Parameters

Estimate	StdErr
87,106	1,609

Estimating the regression parameter and its variance without the weights produces an incorrect estimate of the parameter and an underestimate of the variance.

Example 4.5 Estimating the Variance of a Regression Parameter Estimate Directly

Instead of using PROC SURVEY to correctly estimate the regression parameter estimates and their variances, this example shows how we can do it directly. We do this by calculating parameter estimates for all replicate weights with PROC REG and the sample weights. This produces the correct replicate estimates. Then we apply Step 3 and calculate the squared differences between the estimates. Figure 4.5.5 shows how this can be done.

Figure 4.5.5: SAS Code for Estimating the Variance of a Regression Parameter Directly

```
*This empty data set is produced for the merge later. ;
data data8;
      length parmest0-parmest160 8.;
run;

*Steps 1 & 2: Estimate the replicate estimates for replicates 0 to 160.;
%macro repss(rep);
proc reg data=data1;
```

Figure 4.5.5: SAS Code for Estimating the Variance of a Regression Parameter Directly

```
    model value = bedrms;
    weight repwgt&rep.;
    ods output parameterEstimates = MyParmEst&rep.;
run;
data myparmest&rep.a;
      set myparmest&rep.;
      rename estimate=parmest&rep. StdErr=se&rep.;
      drop model dependent df tvalue probt label;
      if _n_=2 then output;
run;
data data8;
      merge data8 myparmest&rep.a;
run;
%mend repss;
%macro doit;
      %do i=0 %to 160;
            %repss(&i.);
      %end;
%mend doit;
%doit;

*Apply Step 3 to the replicate estimates and estimate the variance.;
data data9 (keep=parmest0 var se);
     set data8 end=eof;
        *Fill array with the replicate means ;
        array parmest{160} parmest1-parmest160 ;
        *Fill array with the squared diffs. ;
        array sdiffsq{160} sdiffsq1-sdiffsq160;
           do j = 1 to 160 ;
             sdiffsq{j} = (parmest{j} - parmest0)**2 ;
           end;
            *Sum the squared diffs. ;
            totdiff = sum(of sdiffsq1-sdiffsq160);
            var = (4/160) * totdiff;
            se = (var)**(0.5);
     output;
run;
proc print data=data9 noobs ;
     format parmest0 var se comma10.0 ;
run;
```

The execution of the SAS code of Figure 4.5.6 generates the output seen in Figure 4.5.1.

Figure 4.5.6: SAS Output for Estimating Regression Parameters Directly

parmest0	var	se
84,952	5,824,204	2,413

The standard error calculated above in Figure 4.5.6 is the same as estimate of Example 4.4 in Figure 4.4.2.

Example 4.6 Estimating the Variance of a Longitudinal Change

Both this example and the next consider statistics that measure longitudinal change. This example will consider an estimate of change that is calculated at the HU level. Example 4.7 will consider a statistic that measures the change between two cross-sectional estimates. We choose these two different longitudinal statistics because their estimated variances are calculated differently.

In this example, we show how to estimate the variance of the proportion of a given binary variable. Specifically, we estimate the proportion of HUs that have a change in their tenure status between 2011 and 2013.

Let $X_{t,k}$ be the tenure status of sample HU k at time t and it has the values of $X_{t,k} = 1$ if owner and $X_{t,k} = 0$ if renter. Let $Y_{t,k}$ be defined as 1 if $X_{t,k} \neq X_{t-1,k}$ and 0 if $X_{t,k} = X_{t-1,k}$. So when $Y_{t,k} = 1$, it indicates that the tenure status of sample HU k changed between t to $t - 1$ -- either from renter to owner or from owner to renter. Now to estimate the proportion of HUs that change their tenure status, we estimate the mean of $Y_{t,k}$ by choosing a common weight for all of the sample HUs observed in t and $t - 1$. For this example, we choose the weight at $t = 2013$, which we refer to as $w_{t=2013,k}$. So the statistic of interest is the proportion of sample units with a tenure switch, i.e.,

$$p_{t=2013} = \frac{\sum_{k \in s_C} w_{t=2013,k} Y_{t=2013,k}}{\sum_{k \in s_C} w_{t=2013,k}},$$

where s_C is the set of all sample HUs that were completed interviews in both 2011 and 2013.

To estimate the variance, we calculate the mean of $Y_{t,k}$ for replicate 0 and replicates 1 to 160 and then apply equation (3.1). Figure 4.6.1 shows how we do this with SAS.

Figure 4.6.1: SAS Code for Estimating a Longitudinal Estimate of Tenure Change

```
* Rename variables and keep the sample HUs in common. ;

data data2011;
      set data2011;
            tenure2011 = tenure;
            status2011 = status;
      rename REPWGT0-REPWGT160 = Yr2011REPWGT0-Yr2011REPWGT160;
      drop status tenure;
run;
data data2013;
      set data2013;
            tenure2013 = tenure;
```

Figure 4.6.1: SAS Code for Estimating a Longitudinal Estimate of Tenure Change

```
            status2013 = status;
      rename REPWGT0-REPWGT160 = Yr2013REPWGT0-Yr2013REPWGT160;
      drop status tenure;
run;
data all1;
      merge data2011(in=a) data2013(in=b);
            by control ;
      * Keep all matches. ;
      if a and b;
run;

data all2;
      set all1;
            if (tenure2011 = '1' and tenure2013 = '1') or
               (tenure2011 = '2' and tenure2013 = '2') then do;
                  Y_tk = 0; tot = 1;
            end;
            else if (tenure2011 = '1' and tenure2013 = '2') or
                    (tenure2011 = '2' and tenure2013 = '1') then do;
                  Y_tk = 1; tot = 1;
            end;
            else delete;
run;

* This empty data set is produced for the merge later. ;
data data10;
      length p0-p160 8.;
run;

* Steps 1 & 2: Estimate the replicate estimates for replicates 0 to 160.;
%macro repsss(rep);
* Estimate the total of Y_tk for each replicate. ;
proc means data= all2 sum noprint ;
      weight yr2013repwgt&rep.;
      var Y_tk;
      output out=yes sum=weightedchange;
run;
proc means data= all2 sum noprint ;
      weight yr2013repwgt&rep.;
      var tot;
      output out=total sum=weightedtotal;
run;

* Estimate the total of all sample HUs considered for each replicate. ;
data all3rep&rep;
      merge yes(in=a) total(in=b);
            p&rep = weightedchange/weightedtotal;
      keep p&rep;
run;
data data10;
      merge data10 all3rep&rep;
run;
%mend repsss;
%macro doit;
      %do i=0 %to 160;
```

Figure 4.6.1: SAS Code for Estimating a Longitudinal Estimate of Tenure Change

```
            %repsss(&i.);
      %end;
%mend doit;
%doit;

* Apply Step 3 to the replicate estimates and estimate the variance. ;
data data11 (keep=p0 var se);
 set data10 end=eof;
  * Fill array with the replicate means ;
  array pro{160} p1-p160 ;
  * Fill array with the squared diffs. ;
  array sdiffsq{160} sdiffsq1-sdiffsq160;
      do j = 1 to 160 ;
        sdiffsq{j} = (pro{j} - p0)**2 ;
      end;
        * Sum the squared diffs. ;
        totdiff = sum(of sdiffsq1-sdiffsq160);
        var = (4/160) * totdiff;
        se = (var)**(0.5);
      output;
run;
proc print data=data11 noobs ;
        var p0 var se;
        format x 8.4 ;
run;
```

The SAS code of Figure 4.6.2 produces the output in Figure 4.6.1.

Figure 4.6.2: SAS Output for Estimating a Longitudinal Estimate of Tenure Change

p0	var	se
0.0499	0.000001	0.0012

Since our statistic of interest is a weighted mean of $Y_{t,k}$, PROC SURVEYMEANS can also be used to estimate the above proportion and its standard error. Figure 4.6.3 shows how this can be done using the SAS data set ALL2 generated in Figure 4.6.2.

Figure 4.6.3: SAS Code for Estimating a Longitudinal Estimate of Tenure Change

```
proc surveymeans data=all2 mean varmethod=brr;
  var Y_tk;
  weight yr2013repwgt0;
  repweights yr2013repwgt1-yr2013repwgt160;
  ods output statistics= mean1;
run;
data mean2;
      set mean1;
            *Remember to multiply standard error by 2;
            se = stderr*2;
            var = se**2;
      keep varname mean se var;
run;
proc print data = mean2;
run;
```

The SAS code of Figure 4.6.3 produces the output in Figure 4.6.4.

Figure 4.6.4: Example of SAS Output

VarName	Mean	se	var
Y_tk	0.049941	.001202009	.000001445

As shown above, the standard errors calculated in Figure 4.6.2 and Figure 4.6.4 are the same.

Example 4.7 Estimating the Variance a Change in Two Rounds of AHS

There are many different statistics that measure the change between two rounds of AHS. Differences, percent change, and ratios can be used to measure how much a given statistic changed from one round of AHS to another. This example considers the housing characteristic of the value of a home and how it can change over time. Both the rate of change and resulting variance are demonstrated. Calculating the variance for these types of statistics has different steps than the calculation of the prior section, although both can be considered longitudinal measures of change.

The statistic of interest is the change in the mean housing cost from 2011 to 2013. Let θ_t be the mean housing cost at time *t*. The statistic of interest is

$$(\Delta\%)_t = \frac{\theta_t - \theta_{t-1}}{\theta_t}.$$

To estimate the variance of $(\Delta\%)_t$, we use the 2011 replicate weights and calculate 160 replicate estimates of $\theta_{t=2011}$ and similarly use the 2013 replicate weights and calculate 160 replicate estimates of $\theta_{t=2013}$. Next, we merge the replicate estimates of $\theta_{t=2011}$ and $\theta_{t=2013}$ by replicate and calculate 160 replicate estimates of $(\Delta\%)_t$. The final step is to apply equation (3.1) to the replicate estimates of $(\Delta\%)_t$.

Figure 4.7.1 shows how this can be done with SAS.

Figure 4.7.1: Example of SAS Code Estimating the Variance of a Rate of Change

```
data data12;
*This empty data set is produced for the merge later.;
      length rate0-rate160 8.;
run;

*Steps 1 & 2: Estimate the replicate estimates for replicates 0-160. ;
%macro repssss(rep);
* Estimate theta for each year. ;
proc means data= data2013 mean noprint ;
      weight Yr2013repwgt&rep.;
      var value;
      where status = '1';
      output out=datam13rep&rep. mean=mean13rep&rep.;
run;
proc means data= data2011 mean noprint ;
      weight Yr2011repwgt&rep.;
      var value;
      where status = '1';
      output out=datam11rep&rep. mean=mean11rep&rep.;
run;

* Merge the replicate estimates of each year by replicate. ;
data datamrep&rep.;
      merge datam13rep&rep. datam11rep&rep.;
            rate&rep. = (mean13rep&rep.-mean11rep&rep.)/mean11rep&rep.;
      keep rate&rep.;
run;
data data12;
      merge data12 datamrep&rep.;
run;
%mend repssss;
%macro doit;
      %do i=0 %to 160;
            %repssss(&i.);
      %end;
%mend doit;
%doit;

* Apply Step 3 to the replicate estimates and estimate the variance. ;
data data13 (keep=rate0 var se);
    set data12 end=eof;
        *Fill array with the replicate means;
        array rate{160} rate1-rate160 ;
```

Figure 4.7.1: Example of SAS Code Estimating the Variance of a Rate of Change

```
        *Fill array with the squared diffs;
        array sdiffsq{160} sdiffsq1-sdiffsq160;
           do j = 1 to 160;
             sdiffsq{j} = (rate{j} - rate0)**2;
           end;
             *Sum the squared diffs. ;
             totdiff = sum(of sdiffsq1-sdiffsq160);
             var = (4/160) * totdiff;
             se = (var)**(0.5);
     output;
run;

proc print data=data13 noobs;
     var rate0 var se;
     format rate0 se 8.4;
run;
```

The SAS code of Figure 4.7.1 produces the output in Figure 4.7.2.

Figure 4.7.2: SAS Output for Estimating the Variance of a Rate of Change

rate0	var	se
0.0357	0.000088	0.0094

5. How to Calculate Confidence Intervals

Lohr (1999; section 9.5) has an excellent review of confidence intervals for survey estimates which we borrow heavily in this section.

Survey estimates that employ sample weights are different than unweighted estimates. However, under certain conditions it can be shown that $\left(\hat{\theta}-\theta\right)\Big/\sqrt{\hat{v}\left(\hat{\theta}\right)}$ asymptotically standard normal for survey estimates (Krewski and Rao 1981). Consequently, when the assumptions are met, an approximate 95% confidence interval for θ may be constructed as

$\hat{\theta} \pm 1.96\sqrt{\hat{v}\left(\hat{\theta}\right)}$, where $\sqrt{\hat{v}\left(\hat{\theta}\right)}$ = Standard Error.

Alternatively, a t_{df} percentile may be substituted for 1.96 with degrees of freedom *df* are equal to the number of replicates. With AHS, we have 160 replicates so *df* = 160 *and* $t_{df=160} = 0.95$.

Example 5.1 Estimating the Confidence Intervals

In this example, we borrow from prior examples and calculate confidence intervals. Table 5.1 summarizes the confidences intervals for several of the previous examples.

Table 5.1: Example of Estimated Confidence Intervals

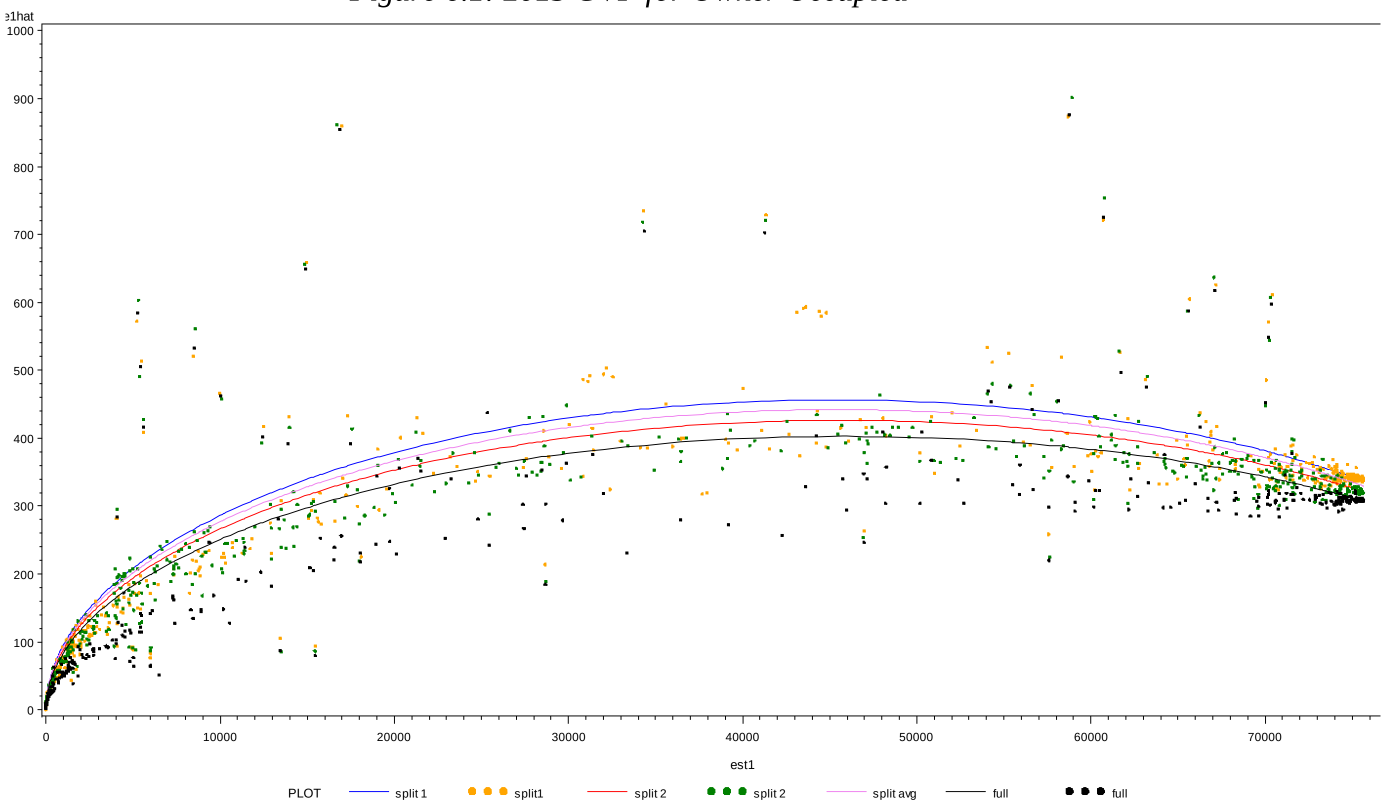
Example	Estimate	Standard Error	Confidence Interval	Coefficient of Variation
2.1 Total with GVF	75,650,326	301,915	(75,058,572, 76,242,079)	0.40%
3.2 Mean number of persons	2.50	0.0064	(2.49, 2.51)	0.26%
4.1 Total with Replication	75,650,326	307,982	(75,046,681, 76,253,971)	0.41%
4.2 Mean	2.5629	0.0073	(2.5486, 2.5772)	0.28%
4.3 Median	160,000	6,519	(147,223, 172,777)	4.07%
4.4 Regression				
–using the sample weights correctly	84,952	2,413	(80,223, 89,681)	2.84%
– using the sample weights for the estimate, but not used correctly with variances	84,952	1,586	(81,843, 88,061)	1.87%
– not using the sample weights at all	87,106	1,609	(83,952, 90,260)	1.85%
4.5 Regression with variance by hand	84,952	2,413	(80,223, 89,681)	2.84%
4.6 Proportion tenure change	4.99%	0.1%	(4.79%, 5.19%)	2.00%
4.7 Percent difference	3.57%	0.94%	(1.73%, 5.41%)	26.33%

We have not covered hypothesis testing. We do point the reader to Schenker and Gentleman (2001) who describe the pitfalls of using confidence intervals in hypothesis testing.

6. How GVFs Are Calculated

To estimate the GVF parameters, the Census Bureau first uses the final weights to estimate the total number of housing units. The totals are estimated for all housing units and by a variety of key domains. Then the replicate weights are used to calculate direct variance estimates for all of the totals. Next, we fit a regression model using (2.1) with the estimated number of housing units, i.e., $\hat{N}$, as the independent variable and the estimated variance of the number of housing units, i.e., $\hat{v}\left(\hat{N}\right)$, as the dependent variable. Figure 6.1 provides an example GVF for owner occupied housing units in 2013.

Figure 6.1: 2013 GVF for Owner Occupied



In Figure 6.1, the horizontal axis represents the range of values for the estimated number of housing units within the Owner Occupied domain. The Owner Occupied domain is the set of all sample HUs where STATUS='1' and TENURE='1'. The vertical axis represents the range of values for the variance of the estimate. Each point in the graph is the combination of an estimate and its variance.

AHS provides multiple GVFs for different domains. Each of the domain-level GVFs are produced using only the domain of interest so it is more suitable for the domain.

As mentioned during the discussion of limitations in section 2, different GVFs are produced for different domains of interest or subsets of the universe of interest and are only appropriate for estimating the variance for those domains of interest.

GVFs for the Split Samples

Since the split samples each include half the sample that the full sample used, we use different weights for the full and split samples. As a result, they have different variances and GVFs. Figure 6.1 shows how the GVFs of split sample 1 and 2 are slightly different. The line between the GVFs of the separate split 1 and split 2 GVFs is the average of the split 1 and 2 GVFs. We found the average GVF by averaging the regression parameters of split sample 1 and 2.

Generally, split sample 1 and split sample 2 have the same sample design so we expect them to have the same variance. This is why we provide only one GVF for both of the split samples.

7. How the Replicate Weights Are Calculated

As with most large-scale household surveys, in an effort to control costs, the AHS uses a complex sample design involving stratification, multi-stage sampling, and unequal sampling rates. Weights are needed in the analysis to compensate for unequal sampling rates as well as for non-response. Further, most estimates from complex samples are non-linear statistics, so estimates of the standard errors are often obtained using the first-order Taylor series approximations or replication methods such as balanced repeated replication or jackknife replication. Therefore, the complex sample design needs to be taken into account in estimating the precision of survey estimates. Not accounting for these sample design features will lead to inaccurate point estimates and an underestimation of the precision. This section describes methods used for generating point estimates and variance estimation.

Variance Estimates with Replication

Replication methods are able to provide estimates of variance for a wide variety of designs using probability sampling, even when complex estimation procedures are used. This method requires that the sample selection, the collection of data, and the estimation procedures be carried out (replicated) several times. The dispersion of the resulting estimates can be used to measure the variance of the full sample.

Two Methods of Variance Replication

In the variance estimation of AHS, we use two types of replicate variance estimation techniques: Balanced Repeated Replication and Successive Differences. Both of the techniques are embodied by the replicate factors that are produced for the AHS replicate variance estimator.

Replicate Weights

The unbiased weights (baseweight × special weighting factor) are multiplied by the replicate factors to produce unbiased replicate weights. These unbiased replicate weights are further adjusted through the noninterview adjustment, the first-stage ratio adjustment, and the second-stage ratio adjustments (including any necessary raking) just as the full sample is weighted. By applying the other weighting adjustments to each replicate, the final replicate weights reflect the impact of the weighting adjustments on the variance.

Replicate Factors for NSR Strata

For the NSR strata, we use replicate factors that are based on the Balanced Repeated Replication (BRR) variance estimator (McCarthy 1966). General reviews of BRR are also found in Wolter (1985; chapter 3) and Särndal et al. (1992; section 11.4). These replicate factors are used to measure the variance due to the selection of the first-stage sample. In SR strata, no PSUs were selected so BRR replicate factors are not appropriate. Since the variation of the SR PSUs comes entirely from selecting units within the PSU, we use the successive difference replication (SDR) method as defined by Fay and Train (1995) for SR PSUs.

To assign the BRR replicate factors, the NSR strata are combined into pseudo-strata, and one NSR PSU from the pseudo stratum is randomly assigned to one of the two half-samples. Fay's method (Fay 1989, Judkins 1990) was used to assigned replicate factors of 1.5 or 0.5. These factors are assigned based on a single row from a Hadamard matrix and are further adjusted to account for the unequal sizes of the original strata within the pseudo stratum (Wolter 1985). For more information about Hadamard matrices and BRR, see Wolter (1985).

These factors were further adjusted to account for the unequal sizes of the original strata within the pseudostratum. All units in a pseudostratum are assigned the same row number(s).

Replicate Factors for SR Strata

The theoretical basis for the successive difference method was discussed by Wolter (1984) and extended by Fay and Train (1995) to produce the SDR method. The following is a description of this method.

To apply SDR to the SR sample, we sort the SR sample by PSU and then within PSU by the same order that was used to select the original *systematic* sample. Then each sample unit is assigned two rows of the given Hadamard matrix. For example, the assignment for a Hadamard of order 160 would be rows (1,2) assigned to the first unit, rows (2,3) assigned to the second unit, … rows (160,1) assigned to the 160th unit. The assignment is repeated in further cycles until the entire sample is assigned two rows.

For an SR sample, two rows of the Hadamard matrix are assigned to each pair of units creating replicate factors, f_r for r = 1,..., R as

$$f_{i,r} = 1 + 2^{-\frac{3}{2}} h_{i+1,r} - 2^{-\frac{3}{2}} h_{i+2,r}$$

where

i	the index on the units of the sample
r	the index on the set of replicates
$h_{i,r}$	number in the Hadamard matrix (+1 or −1) for the *i*th unit in the systematic sample.
R	the number of total replicate samples or simply replicates

This formula yields replicate factors of approximately 1.7, 1.0, or 0.3.

Example 7.1: Successive Difference Replication

The following is a simple example showing the SDR method. The sample contains the n = 5 units and their weights are shown in Table 7.1.

Table 7.1: Sample Weights

Sample HU	Sample Weight
1	15.00
2	23.00
3	19.00
4	16.00

We choose to use the following 4 × 4 Hadamard matrix to define the replicate factors.

$$\mathbf{H}_4 = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}$$

Two consecutive rows of $\mathbf{H}_4$ are assigned to each sample unit as denoted in Table 7.2.

Table 7.2: Assignment of Rows of the Hadamard Matrix

Sample HU	Sample Weight	Row I	Row II
1	15.00	1	2
2	23.00	2	3
3	19.00	3	4
4	16.00	4	1

Plugging these values into our replicate factor formula of equation (3.1) we get

Sample HU 1:

$$f_{1,1} = 1 + 2^{-\frac{3}{2}} h_{1,1} - 2^{-\frac{3}{2}} h_{2,1} = 1 + 2^{-\frac{3}{2}}(+1) - 2^{-\frac{3}{2}}(+1) = 1.0$$

$$f_{1,2} = 1 + 2^{-\frac{3}{2}} h_{1,2} - 2^{-\frac{3}{2}} h_{2,2} = 1 + 2^{-\frac{3}{2}}(+1) - 2^{-\frac{3}{2}}(-1) = 1 + \frac{1}{\sqrt{2}} \cong 1.7$$

$$f_{1,3} = 1 + 2^{-\frac{3}{2}} h_{1,3} - 2^{-\frac{3}{2}} h_{2,3} = 1 + 2^{-\frac{3}{2}}(+1) - 2^{-\frac{3}{2}}(+1) = 1.0$$

$$f_{1,4} = 1 + 2^{-\frac{3}{2}} h_{1,4} - 2^{-\frac{3}{2}} h_{2,4} = 1 + 2^{-\frac{3}{2}}(+1) - 2^{-\frac{3}{2}}(-1) = 1 + \frac{1}{\sqrt{2}} \cong 1.7$$

Sample HU 2:

$$f_{3,1} = 1 + 2^{-\frac{3}{2}} h_{2,1} - 2^{-\frac{3}{2}} h_{3,1} = 1 + 2^{-\frac{3}{2}}(+1) - 2^{-\frac{3}{2}}(+1) = 1.0$$

$$f_{3,2} = 1 + 2^{-\frac{3}{2}} h_{2,2} - 2^{-\frac{3}{2}} h_{3,2} = 1 + 2^{-\frac{3}{2}}(-1) - 2^{-\frac{3}{2}}(+1) = 1 - \frac{1}{\sqrt{2}} \cong 0.3$$

$$f_{3,3} = 1 + 2^{-\frac{3}{2}} h_{2,3} - 2^{-\frac{3}{2}} h_{3,3} = 1 + 2^{-\frac{3}{2}}(+1) - 2^{-\frac{3}{2}}(-1) = 1 + \frac{1}{\sqrt{2}} \cong 1.7$$

$$f_{3,4} = 1 + 2^{-\frac{3}{2}} h_{2,4} - 2^{-\frac{3}{2}} h_{3,4} = 1 + 2^{-\frac{3}{2}}(-1) - 2^{-\frac{3}{2}}(-1) = 1.0$$

Sample HU 3:

$$f_{2,1} = 1 + 2^{-\frac{3}{2}} h_{3,1} - 2^{-\frac{3}{2}} h_{4,1} = 1 + 2^{-\frac{3}{2}}(+1) - 2^{-\frac{3}{2}}(+1) = 1.0$$

$$f_{2,2} = 1 + 2^{-\frac{3}{2}} h_{3,2} - 2^{-\frac{3}{2}} h_{4,2} = 1 + 2^{-\frac{3}{2}}(+1) - 2^{-\frac{3}{2}}(-1) = 1 + \frac{1}{\sqrt{2}} \cong 1.7$$

$$f_{2,3} = 1 + 2^{-\frac{3}{2}} h_{3,3} - 2^{-\frac{3}{2}} h_{4,3} = 1 + 2^{-\frac{3}{2}}(-1) - 2^{-\frac{3}{2}}(-1) = 1.0$$

$$f_{2,4} = 1 + 2^{-\frac{3}{2}} h_{3,4} - 2^{-\frac{3}{2}} h_{4,4} = 1 + 2^{-\frac{3}{2}}(-1) - 2^{-\frac{3}{2}}(+1) = 1 - \frac{1}{\sqrt{2}} \cong 0.3$$

Repeat the process for Sample HU #4 using rows 4 and 1, respectively.

If we calculate the replicate factors for every replicate and unit in the sample, we get the values in Table 7.3.

Table 7.3: Replicate Factors

Sample HU	Replicate Factors			
	Replicate 1	Replicate 2	Replicate 3	Replicate 4
1	1.0	1.7	1.0	1.7
2	1.0	0.3	1.7	1.0
3	1.0	1.7	1.0	0.3
4	1.0	0.3	0.3	1.0

Next multiply the full sample and corresponding factors to get the replicate weights of Table 7.4.

Table 7.4: Final Replicate Weights

Sample HU	Full Sample Weight	Replicate Weights			
		Replicate 1	Replicate 2	Replicate 3	Replicate 4
1	15.0	15.0	25.5	15.0	25.5
2	23.0	23.0	6.9	39.1	23.0
3	19.0	19.0	32.3	19	5.7
4	16.0	16.0	4.8	4.8	16.0

Other Weighting Adjustments for Replicate Weights

In example 7.1, we end at the step of adjusting the replicate base weights for the different replicates. The next step is to calculate the rest of the weighting adjustments for each set of replicate weights. The replicate weights also account for the effect on the variance of the other weighting factors. Recalculating the noninterview and second-stage ratio adjustments for each replicate ensures that the randomness injected or mitigated by the different weighting adjustments is represented in each of the replicate estimates. See also Judkins (1990; p. 224) and Brick and Kalton (1996) for additional discussion of application of other weighting adjustments within replicate weighting.

References

The American Association for Public Opinion Research. 2011. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 7th edition*, AAPOR.

Ash, S. (2014). Using Successive Difference Replication for Estimating Variances," Survey Methodology.

Brick, J.M. and Kalton, G. (1996). "Handling Missing Data in Survey Research," Statistical Methods in Medical Research, 5, 215-238.

Deville, J.C. and Särndal, C.-E. (1992). "Calibration Estimators in Surveys," Journal of the American Statistical Association, 87, 418.

Fay, R.E. (1989). "Theory and Application of Replicate Weighting for Variance Calculations," Joint Statistical Meetings, Proceedings of the Section on Survey Research Methods, 212-219.

Fay, R.E. and Train, G.F. (1995). "Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties," Joint Statistical Meetings, Proceedings of the Section on Government Statistics, 154-159.

Judkins, D.R. (1990). "Fay's Method for Variance Estimation," Journal of Official Statistics, 6, 3, 223-239.

Kreuter, F. and Valliant, R. (2007). "A survey on survey statistics: What is done and can be done in Stata." *The Stata Journal*, 7(1), 1 – 21. Retrieved July 18, 2011 from http://www.stata-journal.com/article.html?article=st0118

Krewski, D. and Rao, J.N.K. (1981). "Inference from Stratified Samples: Properties of the Linearization, Jackknife and Balanced Repeated Replication Method," Annals of Statistics, 9, 5, 1010-1019.

McCarthy, P.J. (1966). "Pseudo-replication: half-samples," Review of the International Statistical Institute, 37, 239-264.

Mukhopadhyay, P.K., An, A.B., Tobias, R.D., and Watts, D.L. (2008). Proceedings of the 2008 NESUG "Try, Try Again: Replication-Based Variance Estimation Methods for Survey Data Analysis in SAS® 9.2".

Plackett, R.L. and Burman, J.P. (1946). "The Design of Optimal Multifactorial Experiments," Biometrika, 33, pp. 305-325.

Särndal, C.E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag.

SAS Institute Inc. 2008. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc.

Schenker, N. and Gentleman, J.F. (2001). "On Judging the Significances of Differences by Examining the Overlap Between Confidence Intervals," The American Statistician, 55:3. August 2001.

Thompson, K.J. and Sigman, R.S. (2000). "Estimation and Replicate Variance Estimation of Median Sales Prices of Sold Houses," Survey Methodology, Vol. 26, No. 2, pp. 153-162.

U.S. Census Bureau. (2006). Current Population Survey: Technical Paper 66.

U.S. Census Bureau, American Housing Survey for the United States: 2009 http://www.census.gov/hhes/www/housing/ahs/nationaldata.html.

Wolter, K.M. (1985). *Introduction to Variance Estimation*, Springer-Verlag.

Glossary

This glossary provides the definitions of several terms used with the technical document. It includes both terms related to statistics and sample design.

1985 Sample Design – refers to the sample design used to select the samples used to make the estimates from 1985 to 2013.

Bias – The formal definition of bias of an estimator $\hat{\theta}$ of some statistic θ is the expected value of the absolute value of the difference between the estimator and statistic and its expected value, i.e., $B(\hat{\theta}) = E\left|\hat{\theta} - \theta\right|$. Informally, the bias is a measure of how close the estimator is to the value it is estimating.

Balanced Repeated Replication (BRR) – A method of variance estimation that is often used with two-stage sample designs that select one or two PSUs per first-stage strata. The method is valuable because it can be applied to estimating the variance of linear and non-linear estimates. Also, the intermediate replicate weights can be provided to data users so that they can estimate variances themselves using simple expressions for the variance. The main ideas of replication are outlined by McCarthy (1966).

Calibration – As described by Deville and Särndal (1992), calibration is a technique that can be used to reduce the variance of an estimator. Sometimes it can also have the effect of improving the coverage of the estimator. Calibration uses a set of known totals: either an "imported totals" (Sarndal and Lundström; p. 54) or a set of variables that are known for all units in the universe. Calibration finds weights that are "close" to the original design weights so that the estimated known totals with the new weights are the same as the known total.

Central City/Balance/Urban/Rural – Geographic identifier which indicates whether the sample block is in a Central City of a 1999 MSA (definition used for the 2000 Census), balance of an (2000 population based) urbanized area, urban cluster, or rural (outside urbanized area or cluster).

Coefficient of Variation (CV) – The square root of the variance of an estimate divided by the estimate, i.e., $\sqrt{v(\hat{\theta})} / \hat{\theta}$.

Core Based Statistical Areas (CBSAs) – A geographic area consisting of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core.

Coverage – A measure of how well a frame and sample design includes the universe of interest. Usually coverage is expressed as a proportion. For example, if a study has 75% coverage then 75% of the universe of interest was included by the frame and the sample design.

Dependent Listing – Starts with a prior list and adds, subtracts, and revises the information on the prior list. See “Independent Listing.”

Domain of Interest or Domain – A specific subset of the universe.

Eligible / Ineligible – Refers to the whether a unit of interest is in the universe of interest or not in the universe of interest. See also AAPOR (2011).

Frame – The list of units in the universe of interest.

Full Sample – Refers to the entire sample and not either of the split samples. Together the two split samples

Generalized Variance Function (GVF) -- As explained by Wolter (1984), the GVF is a simple model that expresses the variance as a function of the expected value of the survey estimate.

Half Sample – Is a code that identifies one of the two PSUs within a pseudo stratum.

Listing – Is the general term for the identification of units. Listing can occur in the geographic area of interest, a building permit office, or within a specific GQ. Once a unit is identified, it is put on a list that can be used as the sampling frame for the survey.

Measure of Size (MOS) – Is a quantity used in unequal sample selection methods to define the probabilities of selection. The MOS is important because sample designs with probabilities of selection proportional to the variable of interest can have small variances. We know this because if the MOS is exactly proportional to the variable of interest, we will have a sampling variance of exactly zero.

Non-interview – Eligible units are either completed interviews or non-interviews. A non-interview occurs for several reasons, such no one was home, household refused to be interviewed, unable to locate the housing unit, or a language problem.

Nonresponse – There are two basic types of nonresponse: unit and item.

Primary Sample Unit (PSU) – The first-stage unit of a multi-stage sample design.

Pseudo Strata – Sometimes called variance strata because they are used in variance estimation or Standard Error Computational Unit (SECU) codes. Since AHS selects one PSU per first-stage stratum, there is unbiased direct estimator of the variance. In order to estimate the variance, we pair the first-stage strata so that we have two PSUs within each strata.

Reference Period – The period of time for which we ask the respondent to report their characteristic of interest.

Relative Variance, Relvariance, or Relvar – is a measure of the relative dispersion of a probability distribution and is defined as the variance divided by the square of the estimate. It is also equal to the square of the coefficient of variation, i.e., $v(\hat{\theta})/\hat{\theta}^2$.

Sample Design – Everything about the selection of units into the sample that determines the probability of selection for each unit. We think of estimation as separate from sample design, in that some estimation procedures are more appropriate than others for a given sample design, but any estimator could be used with the sample derived from a given sample design.

Sampling Fraction – The fraction of the universe that is in the sample. With an equal probability sample design, the sampling fraction is the ratio of the sample size and the size of the universe, often represented as $f = n / N$.

Sampling Interval – The inverse of the sampling fraction. Sometimes referred to as the "take-every" because we take every f^{-1} units of the universe into the sample.

Self-Representing/Non self-representing (SR/NSR) – A unit is self-representing (SR) if its probability of selection is 1.0 and non self-representing (NSR) otherwise. A unit that is SR "represents itself" and no other PSUs because it is the only PSU in its stratum. A unit that is NSR "represents itself and the other units of the same stratum. The terms certainty and non-certainty are used in the same way as SR and NSR, respectively

Self-weighting – Is a type of sample design where units have equal probabilities of selection. AHS has equal overall probabilities of selection obtained by compensating probabilities at different stages. Many household surveys are self-weighting because not much is known about specific households prior to interviewing so they are considered equally important as contributing to the estimate. See also p. 221 in Kish (1965).

Split Samples – In 2013, the national and metro samples were randomly split into two halves, which we refer to as split sample 1 and split sample 2. Each of the split samples was asked additional and separate modules of questions.

Standard Error – Is the square root of the variance, i.e., $se(\hat{\theta}) = \sqrt{v(\hat{\theta})}$.

Stratified Sampling – Is a sample design that partitions the universe of interest into strata and selects an independent sample from each stratum. "If intelligently used, stratification nearly always results in smaller variance for the estimated mean or total than is given by a comparable simple random sample." (Cochran 1977; p. 99)

sys or systematic random sampling from an ordered list.

Successive Difference Replication (SDR) – A replication variance estimation method that mimics the successive difference variance estimator and can be used to estimate the variance from a *sys* sample design. The main ideas of replication are outline by Fay and Train (1995).

Unit – The following definition is from Hájek (1981, p. 4):

"The units making up the population S may be any elements worth studying – persons, families, farms, account items, temperature readings, and so on – and their nature will be irrelevant for theoretical considerations. We shall assume that the units are identifiable by certain labels (tags, names, addresses) and that we have available a frame (list, map) showing how to reach any unit given its label."

For AHS, unit of interest is the housing unit.

Universe of interest – In finite population sampling, the universe of interest, or simply the universe, is the well-defined set of units for which we would like to produce an estimate.

Variance or Sample Variance – Is a measure of the variability of an estimate. With finite population sampling, variance refers to the measure of how the estimate differs if we were to select other samples. Formally, the variance of an estimator $\hat{\theta}$ is the expected value of the squared difference between the estimator $\hat{\theta}$ and its expected value, i.e., $v(\hat{\theta}) = E(\hat{\theta} - E(\theta))^2$.

AHS-N Generalized Variance Function (GVF) Parameters

Table B1: 2013 AHS-N GVF Parameters for the Full Sample

Domain	Total & Total Occupied		Owner Occupied		Renter Occupied	
	a	*b*	*a*	*b*	*a*	*b*
Total Units	-0.000050	6.68	-0.000077	7.03	-0.000073	3.65
New Construction	-0.000942	3.22	-0.001280	3.65	-0.002640	3.00
Mobile Homes	-0.000577	6.32	-0.000654	5.91	-0.000579	4.85
Black	-0.000135	2.70	-0.000361	2.75	-0.000231	2.41
Hispanic	-0.000075	2.87	-0.000303	3.28	-0.000174	2.15
Elderly	-0.000098	3.14	-0.000128	3.20	-0.000313	1.93
Below Poverty	-0.000025	3.50	-0.000077	3.58	-0.000038	2.79
Central City	+0.000180	5.15	+0.000237	4.66	+0.000099	3.08
Suburbs	+0.000164	5.44	+0.000170	4.90	+0.000090	3.21
Outside MSA	+0.001790	21.2	+0.001570	16.6	+0.002430	9.09
Rural	-0.000131	14.5	-0.000174	10.4	-0.000130	5.41
Urban	-0.000016	5.21	-0.000009	4.36	-0.000044	3.13

Table B2: 2013 AHS-N GVF Parameters for Split Samples

Domain	Total & Total Occupied		Owner Occupied		Renter Occupied	
	a	*b*	*a*	*b*	*a*	*b*
Total Units	-0.000074	9.98	-0.000095	8.58	-0.000118	5.56
New Construction	-0.002094	6.68	-0.002885	7.24	-0.005355	6.46
Mobile Homes	-0.000968	9.96	-0.001110	9.54	-0.000988	8.51
Black	-0.000264	4.95	-0.000732	5.36	-0.000474	4.72
Hispanic	-0.000244	5.42	-0.000721	6.21	-0.000474	4.55
Elderly	-0.000196	6.20	-0.000260	6.37	-0.000577	4.24
Below Poverty	-0.000059	5.95	-0.000204	6.46	-0.000087	5.11
Central City	+0.000172	6.88	+0.000221	6.59	+0.000050	5.11
Suburbs	+0.000148	7.05	+0.000142	6.82	+0.000037	5.34
Outside MSA	+0.001840	20.8	+0.001598	17.1	+0.002339	11.0
Rural	-0.000096	14.1	-0.000181	11.4	-0.000110	7.44
Urban	-0.000032	7.03	-0.000036	6.36	-0.000089	5.10

AHS-N Generalized Variance Function (GVF) Parameters

Table B3: 2013 AHS-N GVF Parameters for Vacant HUs

	Full Sample		*Split Sample*	
Domain	*a*	*b*	*a*	*b*
Seasonal Vacant	-0.000191	10.9	-0.000191	10.9
Year-round Vacant	+0.000445	4.87	+0.000445	4.87
For Rent Vacant	+0.000152	2.05	+0.000040	3.23
For Sale Only Vacant	-0.000533	1.88	-0.000752	3.53
Rented or Sold Vacant	+0.000434	3.30	-0.000463	6.19
Occasional Use/URE Vacant	+0.000996	4.98	+0.000449	7.29
Other Vacant	+0.000229	3.47	-0.000219	5.79

Reference for Tables B1-B3: American Housing Survey for the United States: 2013, Current Housing Reports, Appendix D.

Table B4 on the next page provides the PUF variables that define each of the domains of interest used with the 2013 GVFs.

AHS-N Generalized Variance Function (GVF) Parameters

Table B4: 2013 Public Use File definitions to identify the various domain

Domain	Definition
Occupied	STATUS = '1'
Owner	TENURE = '1'
Renter	TENURE in ('2', '3')
Northeast	REGION = '1'
New England	DIV = '1'
Middle Atlantic	DIV = '2'
Midwest	REGION = '2'
East North Central	DIV = '3'
West North Central	DIV = '4'
South	REGION = '3'
South Atlantic	DIV = '5'
East South Central	DIV = '6'
West South Central	DIV = '7'
West	REGION = '4'
Mountain	DIV = '8'
Pacific	DIV = '9'
New Construction	NEWC = '1'
One unit, detached	NUNIT2 = '1'
One unit, attached	NUNIT2 = '2'
Multiunit	NUNIT2 = '3'
Number of units	NUNITS = 2, 3,…
Manufactured / Mobile Homes	NUNIT2 = '4'
Black alone	HHRACE = '02'
Hispanic	HHSPAN = '1'
Elderly	HHAGE >= 65
Below Poverty	POOR < 1000
Central City	METRO3='1'
Suburbs	METRO3 in ('2','3')
Outside MSA	METRO3 in ('4','5')
Rural	METRO3 in ('3','5')
Urban	METRO3 in ('1','2','4')
Seasonal	8 <= VACANCY <= 11
Total Year-round Vacant	1 <= VACANCY <= 7
For Rent	STATUS = '3' and 1 <= VACANCY <= 2
For Sale Only	STATUS = '3' and VACANCY = 3
Rented or Sold	STATUS = '3' and 4 <=VACANCY<= 5
Occasional Use / URE	(STATUS = '3' and VACANCY = 6) or (STATUS = '2' and 1 <=VACANCY<= 7)
Other Vacant	STATUS ='3' and VACANCY=7

AHS-N Generalized Variance Function (GVF) Parameters

Table B5: 2011 AHS-N GVF Parameters

Domain	Total & Total Occupied		Owner Occupied		Renter Occupied	
	a	*b*	*a*	*b*	*a*	*b*
Total Units	-0.000024	4.12	-0.000024	4.12	-0.000047	2.55
New Construction	-0.000096	2.63	-0.000096	2.63	-0.000096	2.63
Mobile Homes	-0.000125	4.24	-0.000125	4.24	-0.000125	4.24
Black	-0.000125	2.46	-0.000319	2.59	-0.000241	2.41
Hispanic	-0.000140	2.30	-0.000321	2.42	-0.000265	2.20
Elderly	-0.000072	2.55	-0.000097	2.65	-0.000344	2.18
Below Poverty	-0.000004	2.84	-0.000004	2.84	-0.000004	2.84
Central City	+0.000383	2.47	+0.000383	2.47	+0.000383	2.47
Suburbs	+0.000196	2.51	+0.000196	2.51	+0.000196	2.51
Outside MSA	+0.002689	3.40	+0.002689	3.40	+0.002689	3.40
Rural	+0.000237	3.32	+0.000237	3.32	+0.000237	3.32
Urban	+0.000043	2.52	+0.000043	2.52	-0.000022	2.35

Table B6: 2011 AHS-N GVF Parameters for Seasonal and Vacant HUs

Domain	*a*	*b*
Seasonal Vacant	+0.001884	4.54
Year-round Vacant	+0.000476	2.69
For Rent Vacant	+0.000154	1.73
For Sale Only Vacant	-0.000546	2.02
Rented or Sold Vacant	+0.001128	2.41
Occasional Use/URE Vacant	+0.002141	2.63
Other Vacant	+0.000110	2.51

AHS-N Generalized Variance Function (GVF) Parameters

Table B7: 2011 AHS-N GVF Parameters for Home Improvement Characteristics

Domain	*a*	*b*
Job Counts	+0.000112	4.12
Job Expenditures	-0.000042	175,000

Reference for Tables B5 – B7: American Housing Survey for the United States: 2011, Current Housing Reports, Appendix D.

Table B8: 2009 AHS-N GVF Parameters

Domain	General Formulas		Other Formulas			
	All characteristics except those listed under *Other Formulas*		Fuels, heating/cooling equipment and neighborhood characteristics		Special Characteristics	
	a	*b*	*a*	*b*	*a*	*b*
Total Units, Midwest, West, Elderly, Black, New Construction, Manufactured/Mobile Homes, Vacant	-0.000026	3.36	-0.000026	3.35	-0.000247	4.09
Northeast, Central City, Hispanic, Urban, Suburbs	-0.000021	2.67	-0.000021	2.67	-0.000247	4.09
Rural, South, Outside (P)MSA	-0.000025	3.21	-0.000032	4.09	-0.000247	4.09
Special Living Sample Units	-0.000012	1.58	-0.000012	1.58	-0.000255	3.85

Reference for Table B8: American Housing Survey for the United States: 2009, Current Housing Reports, Appendix D.

AHS-N Generalized Variance Function (GVF) Parameters

Table B9: 2007 AHS-N GVF Parameters

Domain	General Formulas		Other Formulas			
	All characteristics except those listed under *Other Formulas*		Fuels, heating/cooling equipment and neighborhood characteristics		Special Characteristics	
	a	*b*	*a*	*b*	*a*	*b*
Total Units, Midwest, West, Elderly, Black, New Construction, Manufactured/Mobile Homes, Vacant	-0.000027	3.47	-0.000027	3.47	+0.000255	4.23
Northeast, Central City, Hispanic, Urban, Suburbs	-0.000022	2.67	-0.000022	2.67	+0.000255	4.23
Rural, South, Outside (P)MSA	-0.000026	3.32	-0.000033	4.23	+0.000255	4.23
Special Living Sample Units	-0.000012	1.58	-0.000012	1.58	+0.000255	3.85

Reference for Table B9: American Housing Survey for the United States: 2007, Current Housing Reports, Appendix D.

Table B10: 2005 AHS-N GVF Parameters

Domain	General Formulas		Other Formulas			
	All characteristics except those listed under *Other Formulas*		Fuels, heating/cooling equipment and neighborhood characteristics		Special Characteristics	
	a	*b*	*a*	*b*	*a*	*b*
Total Units, Midwest, West, Elderly, Black, New Construction, Manufactured/Mobile Homes, Vacant	-0.000025	3.16	-0.000025	3.16	+0.000255	3.85
Northeast, Central City, Hispanic, Urban, Suburbs	-0.000020	2.51	-0.000020	2.51	+0.000255	3.85
Rural, South, Outside (P)MSA	-0.000024	3.02	-0.000031	3.85	+0.000255	3.85
Special Living Sample Units	-0.000013	1.58	-0.000013	1.58	+0.000255	3.85

Reference for Table B10: American Housing Survey for the United States: 2005, Current Housing Reports, Appendix D.

AHS-N Generalized Variance Function (GVF) Parameters

Table B11: 2003 AHS-N GVF Parameters

Domain	General Formulas		Other Formulas			
	All characteristics except those listed under *Other Formulas*		Fuels, heating/cooling equipment and neighborhood characteristics		Special Characteristics	
	a	*b*	*a*	*b*	*a*	*b*
Total Units, Midwest, Northeast, West, Elderly, New Construction, Vacant	-0.000021	2.48	-0.000039	4.74	+0.000605	5.53
Central City, Mobile Homes, Hispanic, Urban, Suburbs, Black	-0.000021	2.48	-0.000021	2.48	+0.000924	3.52
Rural, South, Outside (P)MSA	-0.000026	3.12	-0.000039	4.74	+0.000605	5.53

Table B12: 2003 AHS-N GVF Parameters for MSAs

MSA	*a*	*b*
Chicago, IL	-0.000344	1.100
Detroit, MI	-0.000579	1.100
Los Angeles, CA	-0.000332	1.100
New York, NY	-0.000485	2.350
Northern New Jersey	-0.000888	2.300
Philadelphia, PA	-0.000534	1.100

Reference for Table B11-B12: American Housing Survey for the United States: 2003, Current Housing Reports, Appendix D.

AHS-N Generalized Variance Function (GVF) Parameters

Table B13: 2001 AHS-N GVF Parameters

Domain	General Formulas		Other Formulas			
	All characteristics except those listed under *Other Formulas*		Fuels, heating/cooling equipment and neighborhood characteristics		Special Characteristics	
	a	*b*	*a*	*b*	*a*	*b*
Total Units, Midwest, West, Elderly, Black, New Construction, Manufactured/Mobile Homes, Vacants	-0.000027	3.16	-0.000027	3.16	+0.000255	3.85
Northeast, Central City, Hispanic, Urban, Suburbs	-0.000021	2.51	-0.000021	2.51	+0.000255	3.85
Rural, South, Outside (P)MSA	-0.000025	3.02	-0.000032	3.85	+0.000255	3.85

Reference for Table B13: American Housing Survey for the United States: 2001, Current Housing Reports, Appendix D.

Table B14: 1999 AHS-N GVF Parameters

Domain	General Formulas		Other Formulas			
	All characteristics except those listed under *Other Formulas*		Fuels, heating/cooling equipment and neighborhood characteristics		Special Characteristics	
	a	*b*	*a*	*b*	*a*	*b*
Total Units, Elderly, New Construction, Vacants, Northeast, Midwest, West	-0.000022	2.48	-0.000041	4.74	+0.000605	5.53
Central City, Mobile Homes, Hispanic, Urban, Suburbs, Black	-0.000022	2.48	-0.000022	2.48	+0.000924	3.52
Rural, South, Outside (P)MSA	-0.000027	3.12	-0.000041	4.74	+0.000605	5.53

AHS-N Generalized Variance Function (GVF) Parameters

Table B15: 1999 AHS-N GVF Parameters for MSAs

MSA	*a*	*b*
Chicago, IL	-0.000359	1.10
Detroit, MI	-0.000586	1.10
Los Angeles, CA	-0.000336	1.10
New York, NY	-0.000509	2.35
Northern New Jersey	-0.000919	2.30
Philadelphia, PA	-0.000543	1.10

Reference for Table B14-B15: American Housing Survey for the United States: 2003, Current Housing Reports, Appendix D.

Table B16: 1997 AHS-N GVF Parameters

Domain	General Formulas		Other Formulas			
	All characteristics except those listed under *Other Formulas*		Fuels, heating/cooling equipment and neighborhood characteristics		Special Characteristics	
	a	*b*	*a*	*b*	*a*	*b*
Total Units, Midwest, West, Elderly, Black, New Construction, Mobile Homes, Vacants	-0.000028	3.16	-0.000028	3.16	+0.000255	3.85
Northeast, Central City, Hispanic, Urban, Suburbs	-0.000022	2.51	-0.000022	2.51	+0.000255	3.85
Rural, South, Outside (P)MSA	-0.000027	3.02	-0.000034	3.85	+0.000255	3.85

Reference for Table B16: American Housing Survey for the United States: 1997, Current Housing Reports, Appendix D.

AHS-N Generalized Variance Function (GVF) Parameters

Table B17: 1993 AHS-N GVF Parameters

Domain	*a*	*b*
U.S. *(except (s))*, Elderly *(except (s))*, Mobile Homes *(except (h/f,n,s))*, New Construction *(except (s))*, Black *(except (s))*	-0.000030	3.16
Midwest *(except (s))*	-0.000123	3.16
West *(except (s))*	-0.000142	3.16
Central City *(except (s))*, Hispanic *(except (s))*, Urban *(except (s))*, MSA-Suburb *(except (s))*	+0.000171	2.51
Northeast *(except (s))*	-0.000119	2.51
Rural *(except (h/f,n,s))*	-0.000028	3.02
South *(except (h/f,n,s))*	-0.000084	3.02
Outside MSA *(except (h/f,n,s))*, Vacants *(except (h/f,n,s))*	-0.000030	3.23

(h/f)=heating and fueling items, (n)=neighborhood items, (s)=special items

Reference for Table B17: American Housing Survey for the United States: 1993, Current Housing Reports, Appendix D.

AHS-MS Generalized Variance Function (GVF) Parameters

Table C1: 2013 AHS-MS GVF Parameters

		Full Sample		Split Samples	
MSA	Domain	*a*	*b*	*a*	*b*
Baltimore, MD	Owner Occupied	-0.000299	0.300	-0.000584	0.623
	Renter Occupied	-0.000340	0.325	-0.000608	0.656
	Total Units	-0.000254	0.291	-0.000529	0.606
	New Construction	-0.012440	0.350	-0.002710	0.665
Boston, MA	Owner Occupied	-0.000415	0.376	-0.000807	0.804
	Renter Occupied	-0.000539	0.385	-0.000961	0.787
	Total Units	-0.000274	0.330	-0.000596	0.716
	New Construction	-0.020790	0.396	-0.009100	0.773
Chicago, IL	Owner Occupied	+0.000106	0.940	+0.000164	1.968
	Renter Occupied	-0.000133	1.013	+0.000022	2.072
	Total Units	-0.000132	1.335	-0.000059	2.319
	New Construction	-0.014540	1.160	+0.022025	2.492
Detroit, MI	Owner Occupied	+0.000163	0.579	+0.000263	1.225
	Renter Occupied	+0.000296	0.594	+0.000614	1.267
	Total Units	-0.000141	0.902	-0.000036	1.510
	Mobile Homes	+0.003870	0.599	+0.006460	1.379
	New Construction	+0.000027	0.674	+0.051815	1.260
Hartford, CT	Owner Occupied	-0.000353	0.129	-0.000689	0.267
	Renter Occupied	-0.000491	0.121	-0.000924	0.260
	Total Units	-0.000236	0.111	-0.000523	0.245
	New Construction	-0.020480	0.115	-0.012910	0.229
Houston, TX	Owner Occupied	-0.000390	0.807	-0.000790	1.698
	Renter Occupied	-0.000415	0.739	-0.000721	1.562
	Total Units	-0.000293	0.702	-0.000645	1.542
	New Construction	-0.004260	0.750	-0.001695	1.710
Miami, FL	Owner Occupied	-0.000295	0.776	-0.000667	1.632
	Renter Occupied	-0.000290	0.807	-0.000654	1.589
	Total Units	-0.000316	0.780	-0.000658	1.625
	New Construction	-0.030270	0.909	-0.064390	1.894
Minneapolis, MN	Owner Occupied	-0.000352	0.391	-0.000658	0.807
	Renter Occupied	-0.000497	0.368	-0.000844	0.799
	Total Units	-0.000255	0.350	-0.000551	0.758
	New Construction	-0.015330	0.393	-0.001350	0.784
New York, NY	Owner Occupied	-0.000149	1.394	-0.000341	2.843
	Renter Occupied	-0.000259	1.355	-0.000368	2.684
	Total Units	-0.000131	1.377	-0.000140	2.686
	New Construction	-0.000792	1.835	-0.000070	3.825
Northern New Jersey	Owner Occupied	+0.000226	0.848	+0.000212	1.708
	Renter Occupied	-0.000143	0.841	-0.000098	1.758
	Total Units	+0.000052	1.002	+0.000073	1.794

AHS-MS Generalized Variance Function (GVF) Parameters

Table C1: 2013 AHS-MS GVF Parameters

		Full Sample		Split Samples	
MSA	Domain	*a*	*b*	*a*	*b*
	New Construction	-0.000042	0.951	+0.004260	1.970
Oklahoma City	Owner Occupied	-0.000262	0.134	-0.000541	0.279
	Renter Occupied	-0.000285	0.129	-0.000754	0.281
	Total Units	-0.000222	0.122	-0.000495	0.271
	Mobile Homes	-0.005810	0.204	-0.012055	0.420
	New Construction	-0.004790	0.150	-0.002185	0.303
Philadelphia, PA	Owner Occupied	+0.000159	0.654	+0.000144	1.352
	Renter Occupied	-0.000076	0.671	-0.000057	1.368
	Total Units	-0.000097	0.707	-0.000073	1.399
	New Construction	-0.001430	1.202	+0.011200	2.088
Rochester, NY	Owner Occupied	-0.000328	0.124	-0.000625	0.259
	Renter Occupied	-0.000439	0.114	-0.000675	0.249
	Total Units	-0.000238	0.109	-0.000526	0.240
	Mobile Homes	-0.007120	0.115	-0.014495	0.235
	New Construction	-0.006290	0.107	+0.006555	0.226
San Antonio, TX	Owner Occupied	-0.000275	0.198	-0.000540	0.420
	Renter Occupied	-0.000304	0.193	-0.000715	0.442
	Total Units	-0.000210	0.180	-0.000493	0.422
	Mobile Homes	-0.004190	0.222	-0.008420	0.448
	New Construction	-0.003510	0.218	-0.000779	0.487
Seattle, WA	Owner Occupied	-0.000405	0.467	-0.000803	1.011
	Renter Occupied	-0.000481	0.446	-0.000928	0.979
	Total Units	-0.000295	0.438	-0.000667	0.990
	New Construction	-0.006820	0.566	+0.003140	1.119
Tampa, FL	Owner Occupied	-0.000384	0.424	-0.000643	0.871
	Renter Occupied	-0.000467	0.407	-0.000867	0.864
	Total Units	-0.000285	0.388	-0.000593	0.808
	Mobile Homes	-0.003730	0.552	-0.007880	1.109
	New Construction	-0.010070	0.464	-0.009065	0.920
Washington, DC	Owner Occupied	-0.000401	0.740	-0.000760	1.509
	Renter Occupied	-0.000496	0.713	-0.000913	1.472
	Total Units	-0.000294	0.665	-0.000615	1.388
	New Construction	-0.010390	0.795	-0.008025	1.578
Orlando, FL	Owner Occupied	-0.000271	0.253	-0.00512	0.506
	Renter Occupied	-0.000192	0.253	-0.000442	0.502
	Total Units	-0.000250	0.239	-0.000520	0.497
	Mobile Homes	-0.004460	0.263	-0.008980	0.541
	New Construction	+0.006880	0.242	+0.011915	0.490
Las Vegas, NV	Owner Occupied	-0.000277	0.249	-0.000709	0.523
	Renter Occupied	-0.000278	0.226	-0.000518	0.451
	Total Units	-0.000260	0.222	-0.000541	0.462

AHS-MS Generalized Variance Function (GVF) Parameters

Table C1: 2013 AHS-MS GVF Parameters

		Full Sample		Split Samples	
MSA	Domain	*a*	*b*	*a*	*b*
	New Construction	-0.005990	0.257	-0.004645	0.524
Nashville, TN	Owner Occupied	-0.000334	0.203	-0.000690	0.403
	Renter Occupied	-0.000424	0.192	-0.000790	0.382
	Total Units	-0.000281	0.192	-0.000557	0.380
	Mobile Homes	-0.006790	0.229	-0.000390	0.460
	New Construction	-0.003150	0.193	-0.000315	0.381
Austin, TX	Owner Occupied	-0.000401	0.229	-0.000854	0.457
	Renter Occupied	-0.000348	0.211	-0.000824	0.435
	Total Units	-0.000285	0.209	-0.000571	0.419
	Mobile Homes	-0.005370	0.219	-0.011045	0.452
	New Construction	-0.001060	0.220	-0.002057	0.453
Jacksonville, FL	Owner Occupied	-0.000231	0.149	-0.000477	0.298
	Renter Occupied	-0.000357	0.154	-0.000707	0.320
	Total Units	-0.000237	0.144	-0.000490	0.297
	Mobile Homes	-0.003170	0.140	-0.000670	0.291
	New Construction	+0.003630	0.145	-0.002942	0.305
Louisville, KY	Owner Occupied	-0.000301	0.154	-0.000610	0.310
	Renter Occupied	-0.000269	0.158	-0.000774	0.326
	Total Units	-0.000263	0.148	-0.000535	0.301
	Mobile Homes	-0.006020	0.160	-0.012595	0.331
	New Construction	+0.003450	0.159	+0.006253	0.321
Richmond, VA	Owner Occupied	-0.000319	0.146	-0.000599	0.299
	Renter Occupied	-0.000392	0.143	-0.000636	0.296
	Total Units	-0.000256	0.137	-0.000535	0.287
	Mobile Homes	-0.008800	0.133	-0.017435	0.274
	New Construction	-0.000249	0.161	+0.000085	0.303
Tucson, AZ	Owner Occupied	-0.000335	0.129	-0.000622	0.251
	Renter Occupied	-0.000458	0.140	-0.000751	0.270
	Total Units	-0.000266	0.119	-0.000536	0.239
	Mobile Homes	-0.002530	0.114	-0.005095	0.230
	New Construction	-0.010450	0.135	-0.014030	0.265

Reference for Table C1: American Housing Survey for the United States: 2013, Current Housing Reports, Appendix D.

AHS-MS Generalized Variance Function (GVF) Parameters

Table C2: 2011 AHS-MS GVF Parameters

MSA	Domain	*a*	*b*
Anaheim, CA	Owner Occupied	-0.000311	0.328
	Renter Occupied	-0.000272	0.287
	Total Units	-0.000287	0.303
	Mobile Homes	-0.010106	0.318
	New Construction	-0.000267	0.282
Buffalo, NY	Owner Occupied	-0.000300	0.156
	Renter Occupied	-0.000275	0.143
	Total Units	-0.000282	0.147
	Mobile Homes	-0.018737	0.193
	New Construction	-0.000323	0.548
Dallas, TX	Owner Occupied	-0.000248	0.420
	Renter Occupied	-0.000287	0.487
	Total Units	-0.000258	0.437
	Mobile Homes	-0.016116	0.830
	New Construction	-0.000323	0.548
Fort Worth, TX	Owner Occupied	-0.000227	0.195
	Renter Occupied	-0.000204	0.175
	Total Units	-0.000221	0.190
	Mobile Homes	-0.007433	0.288
	New Construction	-0.000204	0.175
Milwaukee, WI	Owner Occupied	-0.000285	0.192
	Renter Occupied	-0.000262	0.177
	Total Units	-0.000277	0.187
	Mobile Homes	-0.059904	0.264
	New Construction	-0.000255	0.172
Phoenix, AZ	Owner Occupied	-0.000274	0.501
	Renter Occupied	-0.000229	0.418
	Total Units	-0.000238	0.434
	Mobile Homes	-0.007123	0.387
	New Construction	-0.000212	0.387
Riverside, CA	Owner Occupied	-0.000291	0.442
	Renter Occupied	-0.000222	0.336
	Total Units	-0.000253	0.383
	Mobile Homes	-0.006101	0.695
	New Construction	-0.000245	0.371
San Diego, CA	Owner Occupied	-0.000263	0.310
	Renter Occupied	-0.000264	0.290
	Total Units	-0.000244	0.290
	Mobile Homes	-0.010140	0.438
	New Construction	-0.000223	0.276
Los Angeles, CA	All Estimates	-0.000247	0.856

AHS-MS Generalized Variance Function (GVF) Parameters

Table C2: 2011 AHS-MS GVF Parameters

MSA	Domain	*a*	*b*
Northern New Jersey	Total Units	-0.000222	0.483
	Mobile Homes	-0.010461	0.725
Birmingham, AL	Total Units	-0.000227	0.114
	Mobile Homes	-0.003920	0.195
Cincinnati, OH	Total Units	-0.000221	0.204
	Mobile Homes	-0.011561	0.348
Cleveland, OH	Total Units	-0.000253	0.243
	Mobile Homes	-0.038296	0.411
Columbus, OH	Total Units	-0.000220	0.176
	Mobile Homes	-0.015876	0.245
Denver, Co	Total Units	-0.000221	0.236
	Mobile Homes	-0.016136	0.284
Indianapolis, IN	Total Units	-0.000263	0.202
	Mobile Homes	-0.017248	0.306
Kansas City, MO	Total Units	-0.000226	0.202
	Mobile Homes	-0.018432	0.332
Memphis, TN	Total Units	-0.000246	0.136
	Mobile Homes	-0.010300	0.231
New Orleans, LA	Total Units	-0.000262	0.143
	Mobile Homes	-0.012429	0.219
Virginia Beach, VA	Total Units	-0.000240	0.167
	Mobile Homes	-0.014934	0.266
Pittsburgh, PA	Total Units	-0.000242	0.268
	Mobile Homes	-0.009201	0.404
Portland, OR	Total Units	-0.000207	0.194
	Mobile Homes	-0.007758	0.267
Providence, RI	Total Units	-0.000190	0.111
	Mobile Homes	-0.028400	0.166
San Francisco, CA	Total Units	-0.000256	0.197
	Mobile Homes	-0.059523	0.250
San Jose, CA	Total Units	-0.000243	0.160
	Mobile Homes	-0.008000	0.160
St. Louis, MO	Total Units	-0.000251	0.314
	Mobile Homes	-0.010392	0.502
Charlotte, NC	Total Units	-0.000216	0.162
	Mobile Homes	-0.006249	0.260
Oakland, CA	Total Units	-0.000241	0.240
	Mobile Homes	-0.024917	0.334
Sacramento, CA	Total Units	-0.000245	0.217
	Mobile Homes	-0.013644	0.322

AHS-MS Generalized Variance Function (GVF) Parameters

Reference for Table C2: American Housing Survey for the United States: 2011, Current Housing Reports, Appendix D.

Table C3: 2004 AHS-MS GVF Parameters

MSA	Domain	*a*	*b*
Atlanta, GA	Mobile Homes	-0.009767	0.660
	All Other Estimates	-0.000254	0.440
Cleveland, OH	Mobile Homes	-0.030314	0.415
	All Other Estimates	-0.000291	0.245
Denver, CO	Mobile Homes	-0.013216	0.265
	All Other Estimates	-0.000237	0.220
Hartford, CT	All Estimates	-0.000274	0.135
Indianapolis, IN	Mobile Homes	-0.010843	0.295
	All Other Estimates	-0.000272	0.195
Memphis, TN	Mobile Homes	-0.014035	0.230
	All Other Estimates	-0.000285	0.135
New Orleans, LA	Mobile Homes	-0.008824	0.245
	All Other Estimates	-0.000300	0.160
Oklahoma City, OK	Mobile Homes	-0.005729	0.230
	All Other Estimates	-0.000301	0.135
Pittsburgh, PA	Mobile Homes	-0.007078	0.400
	All Other Estimates	-0.000262	0.265
Sacramento, CA	Mobile Homes	-0.008825	0.320
	All Other Estimates	-0.000311	0.215
San Antonio, TX	Mobile Homes	-0.005728	0.230
	All Other Estimates	-0.000247	0.150
Seattle, WA	Mobile Homes	-0.006430	0.430
	All Other Estimates	-0.000287	0.290
St. Louis, MO	Mobile Homes	-0.009498	0.495
	All Other Estimates	-0.000285	0.310

Reference for Table C3: American Housing Survey for the United States: 2004, Current Housing Reports, Appendix D.

AHS-MS Generalized Variance Function (GVF) Parameters

Table C4: 1998 AHS-MS GVF Parameters

MSA	Domain	*a*	*b*
Baltimore, MD	Mobile Homes	-0.024994	0.405
	All Other Estimates	-0.000263	0.270
Birmingham, AL	Mobile Homes	-0.005367	0.180
	All Other Estimates	-0.000266	0.105
Boston, MA	Mobile Homes	-0.037632	0.515
	All Other Estimates	-0.000275	0.370
Cincinnati, OH	Mobile Homes	-0.014593	0.315
	All Other Estimates	-0.000286	0.185
Houston, TX	Mobile Homes	-0.009146	0.620
	All Other Estimates	-0.000252	0.390
Minneapolis, MN	Mobile Homes	-0.015859	0.445
	All Other Estimates	-0.000261	0.300
Norfolk, VA	Mobile Homes	-0.010117	0.255
	All Other Estimates	-0.000253	0.160
Oakland, CA	Mobile Homes	-0.019269	0.320
	All Other Estimates	-0.000275	0.230
Providence, RI	Mobile Homes	-0.029162	0.165
	All Other Estimates	-0.000268	0.110
Rochester, NY	Mobile Homes	-0.008538	0.175
	All Other Estimates	-0.000268	0.120
Salt Lake City, UT	Mobile Homes	-0.007892	0.115
	All Other Estimates	-0.000259	0.115
San Francisco, CA	Mobile Homes	-0.039167	0.235
	All Other Estimates	-0.000264	0.185
San Jose, CA	Mobile Homes	-0.006167	0.150
	All Other Estimates	-0.000254	0.150
Tampa, FL	Mobile Homes	-0.001870	0.375
	All Other Estimates	-0.000255	0.290
Washington, DC	Mobile Homes	-0.046204	0.740
	All Other Estimates	-0.000256	0.465

Reference for Table C4: American Housing Survey for the United States: 1998, Current Housing Reports, Appendix D.